Claim Amendments

1.   (Previously Presented) A method comprising:

creating a suffix tree to determine the frequency of phrases within a text corpus;

specifying a set of frequently occurring phrases; and

filtering the set of frequently occurring phrases to determine a set of frequently

occurring and unrecognized phrases as entity name and jargon term candidates.

2.   (Original) The method of claim 1 further comprising:

sorting each phrase of the set of frequently occurring phrases in inverse

lexicographical order prior to filtering the set of frequently occurring phrases.

3.   (Original) The method of claim 1 wherein the text corpus is preprocessed.

4.   (Original) The method of claim 3 wherein the text corpus is text of a human

language.

5.   (Original) The method of claim 4 wherein the human language is Chinese.

6.   (Original) The method of claim 4 wherein filtering the set of frequently

occurring phrases includes comparing a component word of a phrase to a dictionary of

common words and excluding the phrase from the set of entity name and jargon term

candidates if the component word is a common word.

7. (Original) The method of claim 4 further comprising:

reducing the set of entity name and jargon term candidates by applying natural language processing rules.

8. (Previously Presented) The method of claim 7 wherein the natural language processing rules are rules selected from the list consisting of morphological rules, semantic rules, and syntactic rules.

9. (Previously Presented) A machine-readable medium containing instructions which, when executed by a processor, cause the processor to perform a method, the method comprising:

creating a suffix tree to determine the frequency of phrases within a text corpus;

specifying a set of frequently occurring phrases; and

filtering the set of frequently occurring phrases to determine a set of frequently occurring and unrecognized phrases as entity name and jargon term candidates.

10. (Original) The machine-readable medium of claim 9 wherein the method further comprises:

sorting each phrase of the set of frequently occurring phrases in inverse lexicographical order prior to filtering the set of frequently occurring phrases.

11. (Original) The machine-readable medium of claim 9 wherein the text corpus is preprocessed.

12. (Original) The machine-readable medium of claim 11 wherein the text corpus is text of a human language.

13. (Original) The machine-readable medium of claim 12 wherein the human language is Chinese.

14. (Original) The machine-readable medium of claim 12 wherein filtering the set of frequently occurring phrases includes comparing a component word of a phrase to a dictionary of common words and excluding the phrase from the set of entity name and jargon term candidates if the component word is a common word.

15. (Original) The machine-readable medium of claim 12 wherein the method further comprises:

reducing the set of entity name and jargon term candidates by applying natural language processing rules.

16. (Previously Presented) The machine-readable medium of claim 15 wherein the natural language processing rules are rules selected from the list consisting of morphological rules, semantic rules, and syntactic rules.

17. (Previously Presented) A system comprising:

a memory having stored therein executable instructions which when executed

by a processor, cause the processor to perform operations comprising:

creating a suffix tree data structure, the suffix tree data structure storing

phrase frequency data for a text corpus;

using the phrase frequency data to specify a set of frequently occurring

phrases; and

filtering the set of frequently occurring phrases to determine a set of

frequently occurring and unrecognized phrases as entity name and jargon

term candidates ; and

a processor to execute the instructions.

18. (Original) The system of claim 17 wherein the operations further comprise:

sorting each phrase of the set of frequently occurring phrases in inverse

lexicographical order prior to filtering the set of frequently occurring phrases.

19. (Original) The system of claim 17 wherein the text corpus is preprocessed.

20. (Original) The system of claim 19 wherein the text corpus is text of a

human language.

21. (Original) The system of claim 20 wherein the human language is

Chinese.

22. (Original) The system of claim 20 wherein filtering the set of frequently occurring phrases includes comparing a component word of a phrase to a dictionary of common words and excluding the phrase from the set of entity name and jargon term candidates if the component word is a common word.

23. (Original) The system of claim 20 further comprising:

reducing the set of entity name and jargon term candidates by applying natural language processing rules.

24. (Previously Presented) The system of claim 23 wherein the natural language processing rules are rules selected from the list consisting of morphological rules, semantic rules, and syntactic rules.

25. (Previously Presented) The method of claim 1, wherein filtering comprises:

excluding a phrase from the set of frequently occurring phrases, wherein the phrase comprises a sub-phrase that occurs at a higher frequency than the phrase.

26. (Previously Presented) The machine-readable medium of claim 9 wherein filtering comprises:

excluding a phrase from the set of frequently occurring phrases, wherein the phrase comprises a sub-phrase that occurs at a higher frequency than the phrase.

27.   (Previously Presented) The system of claim 17 wherein filtering comprises:

excluding a phrase from the set of frequently occurring phrases, wherein the phrase comprises a sub-phrase that occurs at a higher frequency than the phrase.

28.   (New) The method of claim 1, wherein the filtering comprises:

excluding an embedded phrase from the set of frequently occurring phrases, wherein the embedded phrase is embedded by an embedding phrase that occurs at a similar frequency with the embedded phrase.

29.   (New) The machine-readable medium of claim 9 wherein the filtering comprises:

excluding an embedded phrase from the set of frequently occurring phrases, wherein the embedded phrase is embedded by an embedding phrase that occurs at a similar frequency with the embedded phrase.

30.   (New) The system of claim 17 wherein the filtering comprises:

excluding an embedded phrase from the set of frequently occurring phrases, wherein the embedded phrase is embedded by an embedding phrase that occurs at a similar frequency with the embedded phrase.